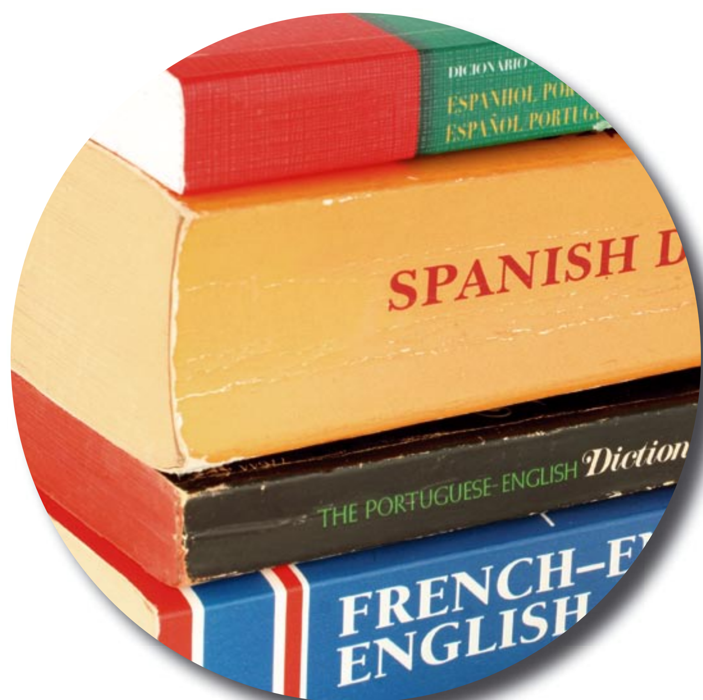
Vertaaldiensten en -programma's

Hoe make you it?

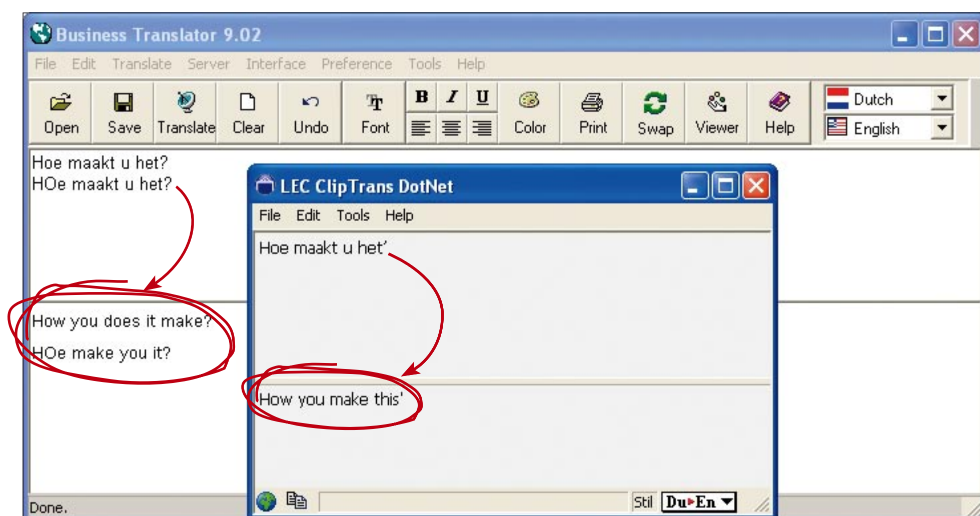
'Hoe maakt u het' klinkt in het Engels als 'How are you' of het iets plechtigere 'How do you do'. Dat is ook wat je zou verwachten als resultaat als je deze simpele zin aan een vertaalprogramma doorspeelt. De praktijk blijkt toch even anders! CEDRIC KNIFE KENS (*)



Het is niet onze gewoonte om onze conclusie al in het begin weg te geven, maar voor dit dossier rond vertaaltools en -services maken we graag een uitzondering. We hebben maar één advies voor wie van plan is om geld neer te leggen voor een 'krachtig' vertaalprogramma dat voor jou gezwind bestanden, teksten en webpagina's vertaalt: doe het niet! We hebben geen enkele commerciële tool of service gevonden die ook maar in de buurt komt van een aanvaardbare vertaling. Niet dat de gratis tegenhangers zoveel beter zijn, maar die hebben wel het voordeel dat ze niets kosten. Commerciële vertaalwoordenboeken die alleen op woordniveau vertalen, zijn over het algemeen wél van goede kwaliteit, maar ook hier zijn er volgens ons genoeg kosteloze diensten en tools die het minstens even goed doen. Een mogelijke uitzondering zijn woordenboeken die zich toeleggen op specifieke semantische velden, zoals wetenschappelijke woordenlijsten.

Losse woorden vertalen is dus geen probleem, maar complete teksten omzetten blijkt zelfs het beste – nu ja, minst slechte – programma moeiteloos op de knieën te krijgen. Vreemd, want bij de eerste experimenten met machinevertaling, nu al meer dan 50 jaar geleden, was men best optimistisch gestemd. Meer zelfs, onderzoekers verwachten nog steeds binnen enkele decennia over uitstekende vertaalcomputers te beschikken.

In dit dossier gaan we eerst na hoe het komt dat (machine)vertalen zo'n lastig karwei blijkt, en vervolgens bekijken we een paar belangrijke evoluties binnen machinevertaling. Daarna laten we enkele concrete vertaaltools en -diensten op je los.



Van 'How you does it make' over 'How you make this' tot 'Hoe make you it'...

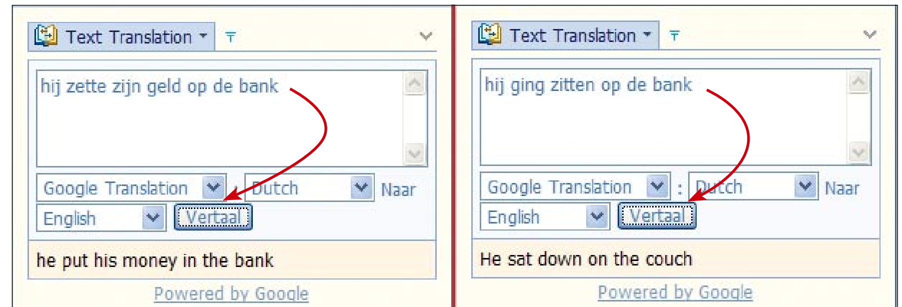
(*) zoals LEC vertaalsoftware Cedric Meskens meende te moeten vertalen.

STRUIKELBLOKKEN

Volgens Ethnologue www.ethnologue.com zijn er momenteel circa 7.000 levende talen bekend, verdeeld over zo'n 108 taalfamilies. Als je weet dat er tussen deze talen – en zeker tussen talen van verschillende families – vaak fundamentele verschillen bestaan qua structuur en grammatica, dan zit je al meteen aan een eerste struikelblok voor machinevertalingen. Zo staan in het Frans de adjectieven bijvoorbeeld vaak na de zelfstandige naamwoorden ('een zeer interessant plan' versus 'un plan très intéressant') en kent het Chinees nauwelijks vervoegingen of meervouden. Het wordt nog veel complexer als we ook de betekenis en de context in beschouwing nemen. Een typisch knelpunt zijn homoniemen, woorden met dezelfde spelling (of klank) maar met een verschillende betekenis. Het Nederlandse woord 'stam' kan je in het Engels onder meer vertalen als 'stem' (woordstam), 'trunk' (boomstam) en 'tribe' (volksstam), en niet zelden valt alleen uit de context af te leiden wat de juiste vertaling is. Om een zin als 'Hij zat op de bank' correct te vertalen als 'He was sitting on the bench' (of 'couch'), moet de vertaalcomputer eigenlijk weten dat je bezwaarlijk op een bankgebouw kan postvatten (dus niet: 'He was sitting on the bank'). Soms volstaat zelfs de tekstuele context niet en moet je ook de socioculturele context of achterliggende communicatieve waarde correct zien in te schatten. Het eenduidige 'Dank u' laat zich in het Japans bijvoorbeeld op diverse manieren vertalen, afhankelijk van de situatie. Maar de makkelijkste manier om vertaalcomputers in hun blootje te zetten, is door ze figuurlijke uitdrukkingen te laten vertalen (tenzij de uitdrukking toevallig in hun databank zit natuurlijk). Je kan dus al raden hoe zowat alle vertaaldiensten 'Hij lacht zich een kriek' omzetten...

VAN REGELSETS NAAR PARALLELTEKST

Machinevertalingen zijn dus allesbehalve evident, en zelfs na meer dan een halve eeuw onderzoek en ontwikkeling blijft het een erg uitdagende technologie. Zowat 40 jaar lang is men blijven focussen op een regelgebaseerde aanpak, waarbij taalkundigen allerlei lexicale en grammaticale regelsets opstelden. Die werden dan in de vorm van algoritmen in de vertaalcomputer geprogrammeerd. Zo'n regel zou bijvoorbeeld kunnen zijn: hou in het Engels onderwerp en persoonsvorm samen, dus niet 'He asked if she a car had', maar wel 'He asked if she had a car'. De redenering was dat computers door het verfijnen van deze regelsets uiteindelijk wel in staat zouden zijn om een degelijke vertaling af te leveren. De praktijk heeft intussen anders bewezen.



Homoniemen: vaak wordt de juiste betekenis pas duidelijk uit de context.

HET SUCCES VAN PARALLEL VERTALEN

Het laatste decennium is machinevertaling echter in een stroomversnelling geraakt. Dat komt niet alleen door de steeds krachtiger wordende computers, maar ook door internationale situaties die vragen om snelle vertalingen – denk bijvoorbeeld aan de VS in Afghanistan. Meer en meer onderzoekers verlaten ook het moeizame pad van de regelsets en richten zich op een statistische aanpak met behulp van parallеltekst. Wat betekent dat? Computers krijgen enorme hoeveelheden brontekst ingelepeld, samen met door deskundigen gemaakte vertalingen. Op basis van die teksten kan je dan tot een nauwkeurigere vertaling komen. De kans is namelijk groot dat een zin als 'How are you today?' verschillende keren (correct) is vertaald als 'Hoe maakt u het vandaag?', zodat de kans kleiner wordt dat die zin in jouw tekst zal opduiken als 'Hoe ben je vandaag?'. Google heeft intussen meer dan 200 miljard woorden uit paralleltekst verwerkt, afkomstig van VN-documenten. Deze vernieuwde aanpak blijkt alvast veelbelovend; niet zelden levert die betere vertalingen op dan via de klassieke weg met regelsets (zie verder). Toch blijft er nog een lange weg te gaan... De nodige parallelteksten vinden mag dan niet zo'n probleem zijn voor wereldtalen, bij kleinere talen wordt dat al een heel stuk lastiger. Bovendien blijken vooral vertalingen van meer zakelijke teksten profijt te halen uit deze methode, ongetwijfeld omdat de gebruikte parallelteksten daar stilistisch en semantisch nauwer bij elkaar aansluiten.

*Paralleltekst levert **betere resultaten** op dan vertalingen op basis van regelsets.*

Blikvanger

Alles-in-één pc Asus bijna klaar

23 oktober Asus' eigen goedkope concurrent van de Apple iMac staat bijna in de winkel. Details over de alles-in-één desktop-pc zijn al uitgelekt. De Taiwanese fabrikant heeft een touchscreendisplay ingebouwd en laat het nieuwe model draaien op een Atom-processor. »

Hints & Tips

Windows Firewall resetten

24 oktober Heb je je pc té goed beveiligd met de Windows Firewall en kan je niet meer surfen? Met een eenvoudig trucje kan je de standaardwaarden weer herstellen. »

Blikvanger

All-in-one PC Asus almost ready

October 23 Asus' own cheap competitor to the Apple iMac is almost at the store. Details on the all-in-one desktop PC have been leaked. The Taiwanese manufacturer has a touch-screen display and built the new model runs on an Atom processor. »

Hints & Tips

Windows Firewall reset

Oct 24 Did you secure your PC with the Windows Firewall, and you can not surf? With a simple trick you can restore the default values.

Blikvanger

Alles-in-één pc. Asus almost ready

23 October Asus' own cheap competitor of Apple iMac stand almost in the shop. Details concerning alles-in-één desktop-pc have already leaked out. The Taiwanese manufacturer has built in touchscreendisplay and lets the new model twist on Atom-processor.

Hints & Tips

Windows Firewall resetten

24 October pc. has protected himself you too well with the Windows Firewall and can you no longer surf? With simple trucje you can repair the defaults.

Links: originele tekst (Clickx-site) – midden: statistische methode met paralleltekst – rechts: regelgebaseerde aanpak.

Online vertaaldiensten

Wie wat rondsnuffelt in Google, botst al snel op heel wat online vertaaldiensten (een zoekterm als *online translation service* doet het 'm wel), maar nogal wat van die services tappen uit hetzelfde vaatje. Heel vaak lees je bijvoorbeeld 'powered by Systran'. Systran is de wereldleider in vertaalsoftware en zweert tot nu toe (overwegend) bij een regelgebaseerde aanpak. Niet zo verwonderlijk dus dat de vertalingen van de meeste vertaaldiensten erg veel overeenkomsten vertonen. Wij beperken ons in dit artikel dan ook tot een drietal grote namen: Babel Fish, Windows Live Translator en Google Vertaling.

Windows Live Translator

Alleen maar Engels

Een relatief jonge loot aan de tak van de vertaaldiensten is Windows Live Translator **www.windowslivetranslator.com**, maar ook deze service tapt uit het alomtegenwoordige Systran-vaatje. De vertaalkwaliteit laat zich dan ook goed vergelijken met die van Yahoo! Babel Fish. We appreciëren wel de realiteitszin van Microsoft wanneer ze schrijven "veel automatisch vertaalde zinnen [zijn] gewoonweg niet te begrijpen". Je kan hier maximaal zo'n 500 à 550 woorden tegelijk invoeren; langere tekst wordt zonder meer afgekapt. Je kan kiezen tussen 25 taalparen, waarin zo'n 12 talen zijn opgenomen. Wat het Nederlands betreft, ben je voorlopig wel beperkt tot omzettingen van en naar het Engels. Om onduidelijke redenen weigerde de dienst af en toe teksten te vertalen en kregen we als enige resultaat de mededeling **Incomplete translation** te zien; over het algemeen nam het vertaalwerk ook iets meer tijd in beslag dan bij Yahoo! Babel Fish. Handig is wel de link **Naar klembord kopiëren**, zodat de vertaling zich makkelijker naar pakweg je tekstverwerker laat overbrengen. Windows Live Translator presenteert de vertaling ook naast de brontekst, wat het makkelijker maakt om zelf de vertaling te checken. Alleen jammer dat de teksten in piepkleine venstertjes verschijnen, waardoor je heel wat moet scrollen.

VERTAALROBOT

Tik je in de onderste balk een webadres, dan krijg je de vertaling in de originele lay-out te zien – voor zover die niet al te complex is, want dan weigert de dienst vaak hele brokken tekst te vertalen. Wel zijn we gecharmeerd door de mogelijkheid om zelf de weergave van de vertaling te kiezen: dat kan naast of onder de oorspronkelijke pagina, maar net zo goed ook in een pop-upvenstertje wanneer je de muispijl boven een zin houdt. Net als bij Yahoo! en Google kan je ook hier met wat knip- en plakwerk (de code voor) een vertaalmodule in je eigen webstek krijgen. Verder biedt de dienst je nog een **Translator**-knop aan voor wie de Windows Live Toolbar heeft geïnstalleerd, zodat je pagina's kan laten vertalen zonder de site te verlaten. En voor wie graag een praatje slaat met een vertaalbot: voeg **mtbot@hotmail.com** toe aan je contactpersonen en ziedaar: alle zinnen die je in Windows Live Messenger intikt, worden meteen vertaald; best leuk als kwaliteit voor jou niet van tel is…



De verschillende weergaves zijn een handig concept – nu de vertaalkwaliteit nog opkrikken…

Babel Fish
900 woorden per keer

Babel Fish is een van de bekendste vertaaldiensten op het net, en ook deze service maakt gebruik van de regelgebaseerde Systran-technologie www.systransoft.be. Tot voor kort kon je hiervoor ook terecht op de website van AltaVista http://babelfish.altavista.com, maar intussen word je automatisch omgeleid naar de stek van Yahoo! http://nl.babelfish.yahoo.com. Jammer, want hoewel de vertaalkwaliteit van beide diensten nagenoeg identiek was, bleek de service van AltaVista guller met tekstlengte. Terwijl Yahoo! je brontekst ongenadig afknipt op zo'n 5.000 à 6.000 tekens (circa 900 woorden) – officieel zelfs al na 150 woorden! – kon je bij AltaVista teksten tot 50.000 tekens en meer kwijt. Voer je hier al te lange teksten in, dan begint Yahoo! Babel Fish zelfs niet aan de vertaling. Nu goed, met 10 keer 5.000 tekens geraak je uiteindelijk ook aan 50.000.

NAAR ENGELS OF FRANS

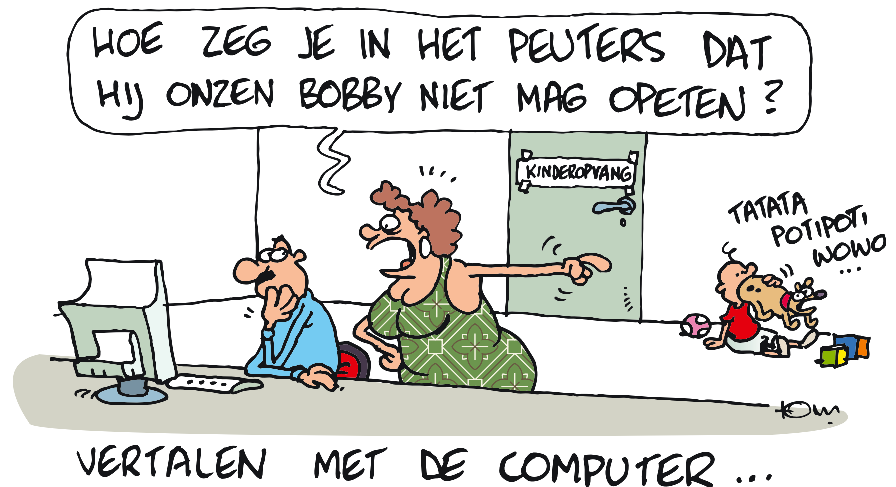
Yahoo! Babel Fish ondersteunt een indrukwekkende 38 taalparen, met daarin een 12-tal verschillende talen. Vertalingen van of naar het Nederlands kunnen echter alleen in combinatie met Frans of Engels. Wil je bijvoorbeeld van het Nederlands naar het Duits vertalen, dan kan je weliswaar Engels als tussentaal gebruiken (Nederlands naar Engels en vervolgens Engels naar Duits), maar de kwaliteit is gezien de dubbele vertaalslag dan vaak nog erbarmelijker. Jammer ook dat de lay-out van de brontekst niet behouden blijft: in de doeltekst komen alle zinnen zonder pardon onder elkaar te staan. Bovendien verschijnt de vertaling in de browser geheel boven je eigen tekst in plaats van ernaast (zoals bij Google), wat controle op de vertaling niet vergemakkelijkt.
Naast eigen tekst (die je zelf intikt of via het Windows klembord inplakt) kan je ook webpagina's laten vertalen: het volstaat om de volledige url in te voeren. Je krijgt het resultaat dan te zien in de originele lay-out van de webpagina. Verder krijg je nog een link aangeboden waarmee je de vertaalservice in de vorm van een bescheiden paneel in je eigen webstek kan onderbrengen. Je hoeft alleen maar een regeltje javascript in de code van je webpagina te kleven.

Google Vertaling
Paralleltekst bij de vleet

Ook Google biedt al langer een vertaalservice aan, maar zowat een jaar geleden gooide de internetmogol het over een andere boeg. Ze zeiden vaarwel aan de regelgebaseerde technologie en schakelden geheel over op een statistische methode met behulp van (massale hoeveelheden) paralleltekst. Als je naar http://translate.google.com surft, krijg je een webpagina met vier tabbladen te zien. Het aanbod op het eerste tabblad, **Tekst en web**, doet sterk denken aan dat van Yahoo! en Microsoft: een kader waarin je zelf tekst kan intikken of plakken en een balkje waarin je de url van een webpagina kwijt kan. In totaal kan je tussen zo'n 34 talen kiezen, waarbij nagenoeg alle onderlinge taalcombinaties mogelijk zijn! Google vermeldt nergens een beperking van het aantal woorden of tekens: teksten met 6.000 woorden (circa 37.000 tekens) werden in elk geval met de glimlach – en sneller dan bij andere diensten – vertaald. Handig is ook dat de vertaling naast de brontekst verschijnt. De kwaliteit is variabel, maar een reeks informele tests gaven ons de stellige indruk dat de vertalingen over het algemeen correcter zijn dan bij Yahoo! of Microsoft.
Vertaalde webpagina's verschijnen ook hier in de oorspronkelijke lay-out, en absoluut handig is dat je in een pop-upvenster de brontekst te zien krijgt wanneer je de muispijl boven een vertaalde zin houdt. Je krijgt telkens ook de gelegenheid om zelf een betere vertaling in te tikken en die via de knop **Bijdragen** richting Google te sturen. In welke mate Google daar effectief rekening mee houdt, is ons niet bekend.

WOORDENBOEKEN

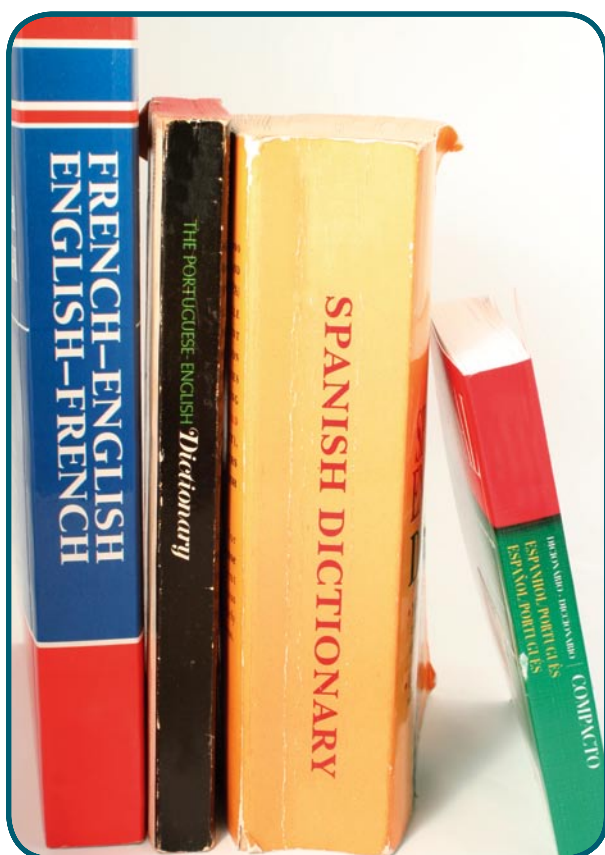
Het tabblad **Woordenboek** houdt weinig verrassingen in: je kan hier terecht voor vertaalwoordenboeken in een dozijn talen, waarbij de bron- of doeltaal wel telkens het Engels is. Het



Nederlands is voorlopig nog karig bemeten: hier geen vertalingen, alleen een verklarend woordenboek waarbij Google gretig put uit de webdefinities van Wikipedia.
Het tabblad **Vertaalde zoekopdracht** blijkt een variant op dit vertaalwoordenboek te bevatten: tik je hier een woord in, dan krijg je aan de rechterkant een reeks websites opgelijst waarin dat woord in de door jou gevraagde doeltaal voorkomt, terwijl je links dezelfde websites ziet maar dan vertaald naar je eigen taal.
Op het tabblad **Hulpprogramma's** ten slotte kan je terecht voor een brokje javascript waarmee je de Google-vertaalmodule in je eigen website integreert. Ook vind je hier een vertaalknop (in de door jou gewenste taal) die je in de werkbalk van je browser kwijt kan. Wanneer je op deze knop drukt, krijg je onmiddellijk de (geselecteerde tekst van de) webpagina in vertaling te zien.

Vertaalprogramma's

Online vertaalservices hebben natuurlijk het nadeel dat je eerst je browser moet opstarten. Dat is niet zo praktisch als je naarstig in je tekstverwerker aan het tikken bent en je de vertaling van een woord of alinea nodig hebt. In dat geval ben je beter af met een vertaalprogramma, dat je net als een andere toepassing lokaal installeert. Zulke programma's vereisen wel een internetverbinding, althans voor vertalingen van tekstfragmenten. Degelijke én gratis programma's liggen niet zo dik bezaaid. Er is wel Babylon **www.babylon.com**, maar deze tool is alleen gratis in een afgeslankte versie met reclame. Ook even opletten als je deze toepassing installeert, want standaard vervangt die de start- en zoekpagina van je browser en trakteert die je tegelijk op een – verwijderbare – toolbar. Laat het echter niet aan je hart komen, want met het (vooralsnog?) gratis Lingoes – 'lingo' betekent vakjargon – kom je even goed aan je trekken, zozeer zelfs dat we er graag een miniworkshop aan wijden... Die vind je op pagina 30. Wel moet je je enkele onhebbelijkheden laten welgevallen. Zo gaat het downloaden van woordenboeken op de site vaak tergend traag – soms zelfs helemaal niet, en dan probeer je best later opnieuw. Bovendien brengt de producent ongeveer om het jaar een vernieuwde versie uit die dan niet (geheel) compatibel blijkt met de vorige, waardoor je de tool en de woordenboeken opnieuw moet installeren.



Toch betalen?

Je kon het al lezen in de inleiding: de door ons getest commerciële tools vinden we vertaaltechnisch niet beter dan de gratis varianten. De eventuele meerwaarde zit in een paar extraatjes.
Het Systran 6.0 Home Pack Belgium (€ 89, onder meer via **www.hcw.be** te verkrijgen) ondersteunt een indrukwekkend aantal taalparen (52 stuks), behoudt bij het vertalen de originele lay-out van je Word-documenten, detecteert net als Google Vertaling automatisch de correcte brontaal en laat je toe eigen woordenboeken te creëren.
LEC Translate DotNet **www.lec.com** biedt dan weer een vertaalprogramma aan in een zestal verschillende gedaantes. Zo zorgt de module **FileTrans** voor een vertaling in batch van bestand(smapp)en, vertaalt **TransIt** gelijktijdig ingetikte tekst en geeft het die door naar een ander venster, bijvoorbeeld van je chatprogramma, en zet **ClipTrans** automatisch alle tekst om die je naar het Windows klembord versast. Het aantal ondersteunde talen mag er best wezen: zo kan je vanuit 20 talen omzetten naar het Nederlands, en laat onze moedertaal zich naar 16 andere talen vertalen. Als we echter de kwaliteit mogen baseren op de vertalingen van en naar het Engels, dan is er nog 'much work on the store'! Je kan het programma een week gratis uittesten, daarna betaal je circa € 4 per maand.

Conclusie

Machinevertaling is al een heel eind gekomen, maar het heeft nog een veel langere weg te gaan... De nieuwe aanpak met statistische analyse van massale hoeveelheden paralleltekst ziet er hoopgevend uit en lijkt de 'gevestigde' aanpak met complexe regelsets te overschaduwen. Hoe dan ook, op dit moment zijn geautomatiseerde tekstvertalingen allesbehalve vlekkeloos en dus niet geschikt om in pakweg zakelijke correspondentie te gebruiken zonder grondige controle door ervaren mensen. Zijn vertaaldiensten dan compleet nutteloos? Alvast niet als het je bedoeling is om een idee te krijgen over de inhoud van de anderstalige brontekst. Tenslotte is zelfs miserabel Nederlands nog altijd beter dan中国赏解, toch?

VAKTAAL

JAVASCRIPT: Eenvoudige scripttaal, losjes gebaseerd op java, waarmee je eenvoudige functies in een webpagina kan programmeren, zoals een klok. Een moderne browser volstaat om javascriptjes te bekijken.

Miniworkshop Lingoes Translator

STAP 1 / TEKST VERTALEN

Surf naar www.lingoes.net en download er het programma Lingoes Translator. Start na de installatie het programma op: een nieuw icoontje verschijnt in het systeemvak van je Windows taakbalk. Wanneer je hierop dubbelklikt, krijg je het hoofdvenster van Lingoes te zien. Klik hier op de knop VERTALING: je krijgt een tekstvak te zien waarin je tekst kan tikken of plakken. Onderaan kan je uit verschillende vertaaldiensten kiezen – waaronder Google Translation, Yahoo! BabelFish en LEC Translation. Afhankelijk van de gekozen vertaaldienst worden ook andere taalcombinaties beschikbaar. Zodra je bron- en doeltaal hebt gekozen, druk je op de knop VERTAAL: de vertaling verschijnt nu in het venster. Leuk gadget voor anglofielen: met het luidsprekericoontje beluister je geselecteerde – Engelstalige – tekstfragmenten.

STAP 2 / 'PRIKBORD VERTALER' GEBRUIKEN

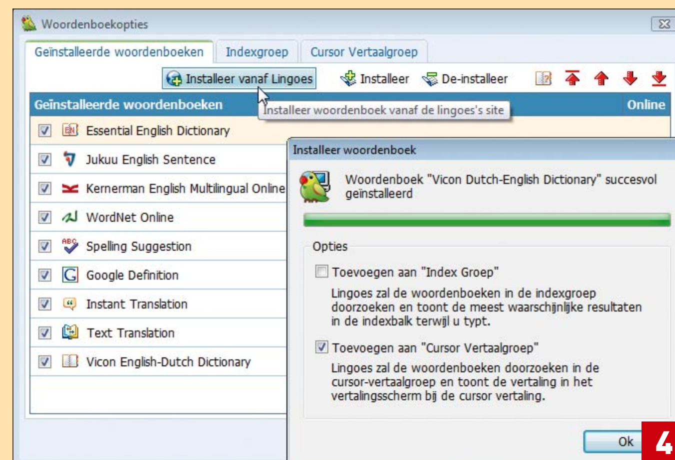
Toegegeven, de aanpak uit stap 1 is behoorlijk omslachtig. Dat moet beter kunnen... Klik met de rechtermuisknop op het Lingoes-icoontje en plaats een vinkje bij PRIKBORD VERTALER. Open nu je favoriete programma en selecteer een willekeurig tekstfragment. Kopieer dit (met de sneltoetscombinatie CTRL+C) naar het Windows klembord: een pop-upvenstertje verschijnt nu met de vertaling! Standaard is dat van het Nederlands naar het Engels, maar andere combinaties zijn ook mogelijk. Open hiervoor nogmaals het contextmenu van Lingoes en kies CONFIGURATIE. Open het tabblad VERTAAL en kies de gewenste taal bij TARGET LANGUAGE (keuze uit 24 talen). Bevestig met OK: voortaan krijg je de vertaling meteen in de geselecteerde doeltaal.

STAP 3 / WOORDEN VERTALEN

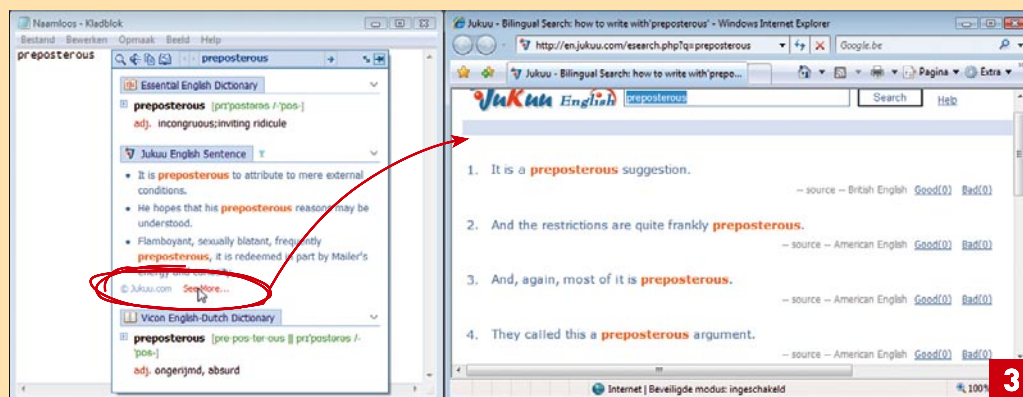
Ben je op zoek naar de vertaling van losse woorden, dan is zelfs de methode uit stap 2 nog wat complex. Geen nood... het kan nog eenvoudiger! Open opnieuw het configuratievenster van Lingoes en ga naar het tabblad TEXT CAPTURE. Standaard staat de MUIS ACTIVERINGSMODUS ingesteld op CTRL+RECHTERMUISKNOP, maar je kan ook een eigen sneltoetscombinatie kiezen. Sluit af met OK. Tik nu bij wijze van experiment een Engelstalig woord in, hou de muiscursor boven dit woord en gebruik de bewuste sneltoetscombinatie: in een venster verschijnen nu Engelstalige definities van die term, inclusief voorbeeldzinnen (zie afbeelding 3). Maar je had hier natuurlijk graag (ook) de Nederlandse vertaling van gezien...

STAP 4 / EXTRA WOORDENBOEKEN RAADPLEGEN

Open alweer het configuratievenster en selecteer dit keer WOORDENBOEKEN. Open het tabblad GEÏNSTALLEERDE WOORDENBOEKEN en druk op de knop INSTALLEER VANAF LINGOES (zie afbeelding 4). Het aanbod is om te watertanden, met gratis kwaliteitswoordenboeken van onder andere Longman, Cambridge en Merriam-Webster. Wij zijn echter op zoek naar vertaalwoordenboeken van en naar het Nederlands: even scrollen en we botsen in de rubriek DUTCH DICTIONARIES op een schitterend duo van Vicon. Klik hier bijvoorbeeld op de link naar VICON ENGLISH-DUTCH DICTIONARY en kies OPENEN. Het wordt meteen geïnstalleerd, waarbij je zeker een vinkje plaatst naast TOEVOEGEN AAN "CURSOR VERTAALGROEP". Voortaan krijg je ook de vertaling te zien als je een Engels woord aanklikt. Nog meer talen? Even speuren op www.lingoes.net dan maar! ♦



Keuze uit tientallen woordenboeken, in tal van talen...



Woordenboeken met een muisklik!